Co-funded by the Erasmus+ Programme of the European Union

Development of the interdisciplinary master program on Computational Linguistics at Central Asian universities (CLASS)

Analyzing the content of international educational master programs in Computational Linguistics

Report

Altynbek Sharipbay (L.N. Gumilyov ENU, CLASS coordinator)

Astana, 2018

L.N. Gumilyov Eurasian National University, Department of Computer Sciense and Information Security

Contact : L.N. Gumilyov ENU, CLASS Coordinator, Prof. A.Sharipbay, +77477771652, sharalt3@gmail.com

Introduction

The information explosion of recent years has significantly changed not only public, but also scientific life: it gave rise to completely new tasks, which in many respects changed the structure of scientific knowledge, in particular, firmly combining its humanitarian component with the field of natural and exact sciences. So a new interdisciplinary direction arose, which is now commonly called Computational Linguistics; it is successfully and dynamically developing as an independent area, that demands competent professionals. Computational Linguistics is a young but very promising science. And the main traits of this science is that it is inherently interdisciplinary, that is, it studies the problems, crossing subjects of computer science and linguistics. And in order to become a full-fledged expert in this field, you need to understand computer science and linguistics.

In the modern information society, the amount of available linguistic data, relating to a variety of areas, has grown exponentially. At the same time, new types of discourse were formed, and the old ones acquired a different status and characteristics. In addition to oral and written versions of the language, Internet communication arose, and the colloquial norm penetrated the written digital space, and globalization caused a qualitatively new round of inter-language interference. In view of this, in a wide variety of social and scientific fields, there was an acute need for theoretical and applied linguists, capable of structuring, analyzing, and processing large amounts of this kind of data. Cooperating with professionals of different subject areas, at the intersection of different specialties, in mixed teams, they create and use new tools to solve complex information problems - labeled text corpora, text databases, electronic dictionaries and thesauruses, interactive and crowdsourcing language resources of various types, etc.

Modern linguistics uses statistical and probabilistic models, and also relies heavily on methods of processing of large data, primarily corpora. Methods of processing of unstructured text information, borrowed from Computational Linguistics, are used not only to study language structures, but also to model humanitarian knowledge (history and theory of literature), based on text analysis. Thus, the statements made in the theoretical aspects should, in the modern understanding, be confirmed by statistical processing of real data. This approach requires not only the possession of a complex of skills related to the processing of large volumes of data, but also a fundamentally new understanding of what is an evidence in linguistics. Without a doubt, the spread of algorithms for statistical processing of text data requires the possession of appropriate computer software tools.

At the same time, demand for specialists who have completed basic training in mathematics and computer science in combination with linguistic training, based on modern theories on the principles of organizing a natural language, is quite high.

Modern Computational Linguistics deals with problems, related to automatic analysis of natural language. This includes well-known areas of machine translation, information retrieval and understanding, voice interface. This also includes creation of teaching and reference language resources, experimental tasks in the field of language theory. This, finally, includes developing of variety applications, that rely on language data. Linguists are in demand in start-ups, developing new linguistic technologies - for example, for creating robots with

natural language interface, or for automatically recognizing emotions in the texts of users of social networks; in companies that do not deal with linguistics, but who need a professional to handle large amounts of unstructured text data.

At the moment, specialists in Computational Linguistics are also in high demand in the largest companies engaged in actual developments in the field of linguistics. For the education in the countries of Central Asia the task of training such personnel is new, it became evident only in the most recent time in connection with the rapid development of this field, and programs for targeted training of computer linguists in the magistracy are almost non-existent, so the educational program on Computational Linguistics is certainly relevant.

In Kazakhstan, in the classifier of specialties for higher education there is no specialty "Computational Linguistics", but at the ENU them. LN Gumilyov since 2013 developed and implemented the trajectory of training "Computational Linguistics" in the framework of educational programs of specialties "5B060200-Informatics" for undergraduate and "6M060200-Informatics" for magistracy. For this purpose, we studied the educational programs on Computational Linguistics of foreign universities.

Review of educational programs on Computational Linguistics of foreign higher educational institutions.

Training of national cadres in the field of Computational Linguistics should be carried out through qualitative improvement of educational programs on the basis of studying international experience in the development and implementation of educational programs in Computational Linguistics, tuning methodology tools, as well as carrying out a questionnaire, diagnosis and classification of training needs.

To develop the Erasmus + Development of the interdisciplinary master program on Computational Linguistics at Central Asian universities (CLASS), we analyzed the educational programs of the following universities and institutes, which are listed in Table 1.

Table 1. Educational programs in Computational Linguistics

№	Country, city, name of the university, Web site	Name of the educational program	Description of the university
1	Russia, Moscow, Russian State University for the Humanities, www.rggu.ru	45.04.03 Fundamental and applied linguistics: Computational Linguistics; Theory of Language	Russian State Humanitarian University (RGGU) is a federal state budget educational institution of higher education. The RSUH conducts the training of students on 39 professional baccalaureate educational programs [8] and 28 directions of the magistracy [9], including: document management and archival studies, international relations, oriental studies and African studies, culturology, religious studies, history, sociology, political science, art history, management, advertising and public relations, economics, philosophy, philology, psychology, intellectual systems in the humanitarian sphere, journalism. More than 10 thousand students are studying at the Russian State University for the Humanities and about 20 thousand - in the branches. The faculty of the university has about 600 full-time teachers and about 200 part-timers, specialists from the RAS institutions, Moscow universities and other scientific institutions. The RSUH employs more than 70 academicians and corresponding members of Russian and foreign academies, more than 200 professors and doctors, and over 500 candidates of science [10]. About 300 graduate students in 36 specialties are trained in the postgraduate course of the RSUH, there is a doctoral program. There are 11 specialized dissertational councils, five of them are on defense of doctoral dissertations. The RGGU has 10 educational and scientific institutes [11]: Historical Archival Institute Institute of Economics, Management and Law Institute of Psychology. L. S. Vygotsky Institute of Mass Media Institute of Linguistics Institute of Information and Security Technologies Institute of Philology and History Institute of Oriental Cultures and Antiquity Institute of New Educational Technologies Russian Anthropological School
2	Russia, St. Petersburg, St. Petersburg State University, Spbu.ru	Engineering of Humanities	St. Petersburg State University is the first university in Russia. It was founded by the decree of Emperor Peter I in 1724. Today St. Petersburg State University trains specialists in almost all the most popular fields of science in the field of specialties. For this, there are 19 faculties, 13 research institutes and more than 4 thousand teachers. In 2017/18, 398 basic educational programs are being implemented in St. Petersburg State University, including 12 general education programs, 11 secondary professional education programs, 102 undergraduate and graduate programs, 181 master's programs, 64 postgraduate programs in scientific specialties and 28 residency programs.
		45.04.02 Applied and experimental linguistics	

3	Russia, Moscow National Research University High School of Economics, www.hse.ru	Computational Linguistics	The Federal State Autonomous Educational Institution of Higher Education "National Research University" Higher School of Economics "was established by the Decree of the Government of Russia on November 27, 1992, initially as a center for the preparation of masters. The Higher School of Economics is a university preparing economists, sociologists, managers and lawyers and leading an active international scientific research activity. Here they teach philosophy, mathematics, history of literature, journalism, psychology, sociology and even design The "QS - World University Rankings by Subject", the subject list "Linguistics" The "QS - World University Rankings by Subject", the subject list "Linguistics" 151-200
4	Russia, Moscow, Moscow University of Physics and Technology MPhTI, mipt.ru	Computational Linguistics	The Computational Linguistics Department of the FIVT was founded in 2011 by the Russian company ABBYY, one of the leading software developers in the field of artificial intelligence, in particular, document recognition and natural language processing. The department trains specialists who can effectively work in the field of development of innovative language computer technologies, in particular, the technology of syntactic and semantic analysisABBYY Compreno
5	Russia, Saint-Petersburg ITMO, **ifmo.ru**	09.04.02 Information systems and technologies Directions: Speech information systems	ITMO University (1900) is a major state university in St. Petersburg, one of the national research universities in Russia. ITMO is among the 15 Russian universities participating in the Project 5-100. The university's scientific priorities are information and photon technologies. ITMO is considered the world's forge of personnel for the IT industry. In 2016, students were enrolled in more than 120 master's programs in 35 areas of training in information and optical technologies, economics, environmental protection, metrology, telecommunications, photonics, and technical physics.
6	USA, Los Angeles (UCLA) University of California, www.linguistics.ucla.edu	Department of Linguistics	The UCLA Linguistics Department focuses on the scientific study of language in all aspects. The fields represented include phonetics, phonology, syntax, and semantics, as well as the interdisciplinary areas of psycholinguistics, language acquisition, historical linguistics, and mathematical linguistics. The department has a strong emphasis on linguistic theory as well as on fieldwork and experimental study.
7	USA, Cambridge, MA, Harvard John A. Paulson School of Engineering and Applied Sciences (SEAS). https://www.seas.harvard.edu	CS50: Introduction to Computer Science I CS51: Introduction to Computer Science II CS182: Artificial Intelligence CS187: Computational Linguistics	SEAS is a department at famous Harvard University and was founded in 1847 (as Lawrence Scientific School). The school has six undergraduate concentrations (Applied Mathematics, Biomedical Engineering, Computer Science, Electrical Engineering, Engineering Sciences, and Mechanical Engineering) and offers a Bachelor of Science, Bachelor of Arts, Master of Science, Master of Engineering, and Doctor of Philosophy
8	UK, Cambridge University, mi.eng.cam.ac.uk/mi/Main/Speech	The Machine Intelligence Laboratory	The Speech Research Group is part of the Machine Intelligence Laboratory of Information Engineering Divisionad Cambridge University. Its primary specialism is in large vocabulary speech transcription and related technologies. It also has active research interests in spoken dialogue systems, multimedia document retrieval, statistical machine translation, speech synthesis and machine learning.

9	USA, Johns Hopkins University, www.clsp.jhu.edu	The Center for Language and Speech Processing	The Johns Hopkins Center for Language and Speech Processing (CLSP) is an interdisciplinary research and educational center focused on the science and technology of language and speech. Within its field, CLSP is recognized as one of the largest and most influential academic research centers in the world.
10	USA, Carnegie Mellon University, www.hcii.cmu.edu	The Human-Computer Interaction Institute (HCII) at Carnegie Mellon School of Computer Science	The Carnegie Mellon School of Computer Science is recognized around the world as a leader in all facets of computer science and robotics education. Respected international surveys have consistently ranked CMU's graduate programs in computer science among the best in the United States. SCS also is ranked highly in specialty areas such as programming languages, artificial intelligence, systems and theory.
	USA, Stanford, CA, Stanford University, https://www.stanford.edu	Master of Arts and a Ph.D in Linguistics.	Stanford University has one of the nation's oldest computational linguistics programs. The Department of Linguistics along with the Computer Science Department hosts an array of research labs and informal workshops on computational linguistics. Stanford University is tied for #1 in computer science.
	USA, Boulder, CO, University of Colorado Boulder www.colorado.edu	Computational Linguistics - Master of Science (MS)	As a comprehensive university, CU Boulder is committed to the liberal education of students via a broad curriculum ranging from the baccalaureate through the postdoctoral levels. The chancellor is the chief academic and administrative officer and is responsible for conducting campus affairs in accordance with the policies of the regents. With an enrollment of more than 32,000 students, the CU Boulder is the largest of the four-campus University of Colorado system. The student population comes from every state in the nation and from more than 130 foreign countries. Many different ethnic, religious, academic and social backgrounds are represented, fostering the development of a multicultural academic community that enriches each student's educational experience.

From this it is clear that although their educational programs of some universities are related to Computational Linguistics in content, they are called in different ways

Results of the questionnaire

In order to know the opinion of specialists in the field of Computational Linguistics, we conducted a survey using the following application: https://docs.google.com/forms/d/e/1FAIpQLScT6iJA5e-iYKIC4mAIzcF_sMch77CRylhGUwmi_FQ1afiRWA/viewform?c=0&w=1&usp=mail_form_link

The purpose of the questionnaire is to determine the place of work (industry) and the competence of specialists in the field of Computational Linguistics based on interviewing potential employers.

The following results of the questionnaire were received:

1. Where, in your opinion, can specialists in the field of Computational Linguistics work?
57,1 % - Publishing companies;
71,4 % - Telecommunications;
75% - educational institutions;
42,9 % - State and administrative;
28,6 % - Small and medium-sized production service companies;

32,61 % - In all companies where there are computers;
75 % - Research laboratories;
28,6% - Public non-governmental organizations;
17,9 % - Other

2. Where, in your opinion, is the greatest need for specialists in Computational Linguistics?
60,7% - Telecommunications;
46,4 % - Publishing companies;
14,3 % - Banks;
7,1 % - Industrial enterprises;
42,9 % - State administrative;
17,9 % - Small and medium-sized production service companies;
25 % - Small and medium-sized enterprises producing products.
35,7 % - Large companies;
28,6% - In all companies where there are computers;
75 % - Research institutes, universities;
7,1 % - Other

3. Specify the types of activities, the main tasks that these specialists should know and implement?
39,3 % - Design;
28,6 % - Production and technology;
32,1 % - Organizational and management;
75% - Research;
64,3 % - Innovative;
50 % - Service-operating;
21,4 % - Other

4. Indicate what problems you encounter in your enterprise regarding Computational Linguistics?
25 % - The problem of confidentiality of information;
14,3 % - The problem of changing data;
17,9 % - The problem of the replacement of personal data;
28,6 % - Ignorance of the modern electronic language environment;
46,4 % - Inability to navigate in computer tools of linguistic analysis and in Computational Linguistics;
35,7 % - Ignorance of the main approaches to digital modeling of humanitarian data in the field of history and literature;
50 % - Inability to design specialized linguistic databases;
35,7 % - Ignorance of various methods of mathematical generalization of the results of linguistic research;
50 % - Inability to conduct processing of linguistic data by modern means;
35,7 % - Ignorance of the principles of constructing various linguistic resources, including the corpus of texts;
42,9 % - Inability to use existing tools and linguistic resources for product development.

5. Indicate what tools (tools) you use in your enterprise for information processing (software and materials)?
60,7 % - Packages of applied programs;
82,1 % - Information retrieval systems;

39,3 % - Software for character processing;
50 % - Office programs;
10,7 % - Other.

6. In your opinion, what general competencies should a specialist in Computational Linguistics have?
53,6 % - Creativity;
39,3 % - Efficiency;
28,6 % - Care for quality;
35,76 % - Work in a team;
39,3 % - the ability to perform and develop their intellectual and general cultural level;
42,9 % - The ability to perceive mathematical, natural science, socioeconomic and professional;
42,9 % - Ability to develop and study techniques for analysis, synthesis, optimization and forecasting;
25 % - Ability to simulate processes and objects based on standard automation packages;
17,9 % - Initiative and ability to take necessary actions;
46,4 % -critical thinking;
39,3 % - Ability to analyze and draw conclusions;
28,6 % - Ability to communicate with a diverse audience;
10,7 % - Confidentiality;
21,4 % - Project design and management;
35,7 % - Ability to work in an international context;
3,6 % - Other.

7. In your opinion, what specific competencies should a specialist in Computational Linguistics have?
75 % - Ability to apply existing tools and linguistic resources for development;
78,6 % - Knowledge of basic levels of analysis and synthesis of text in natural language;
71,4 % - Knowledge of the principles of the construction of various linguistic resources, including the corpus of texts;
39,3 % - Understanding the essential differences of natural languages from the artificial and the feature of Computational Linguistics;
60,7 % - Ability to conduct processing of linguistic data by modern means;
50 % - Knowledge of the understanding of the algorithms of primary processes for automatic processing of text and speech;
39,3 % - Ability to understand which language tools are behind a particular local task;
42,9 % - Ability to use statistical methods of analyzing language data and visualization tools;
50 % - Knowledge of various methods of mathematical communication of the results of linguistic research;
35,7 % - Ability to design specialized linguistic databases;
39,3 % - The ability to correctly use the results of mathematical generalization and use the results obtained;
25 % - Knowledge of basic approaches to digital modeling of humanitarian data in the field of history and literature;
32,1 % - Ability to navigate in computer tools of linguistic analysis;
21,4 % - Ability to evaluate the complexity of different solutions and the thresholds of acceptable solutions;
28,6 % - Ability to program prototypes and decision models;
21,4 % - Ability to design a chain of processing of language data and interpret results;

32,1 % - Knowledge of database and service management systems (SQL Server);
39,3 % - Ability to understand the current state of Computational Linguistics and information;
39,3 % - Ability to conduct independent research and obtain new scientific results in the field of Computational Linguistics;
39,3 % - Work in interdisciplinary work, ability to interact with experts in other subjects;
21,4 % - Solving a wide range of known theoretical and practical problems of computer facilities.

8. In your opinion, what subjects will allow these specific competences to be obtained?
64,3 % - English language;
46,4 % - Kazakh language;
60,7 % - Digital signal processing;
64,3 % - Automatic text recognition;
28,6 % - Recognition of announcers;
64,3 % - Syntactic analysis of texts;
71,4 % - Morphological analysis of texts;
57,1 % - Methods of transcription of sounds and speech;
60,7 % - Methods of speech synthesis;
50 % - Specialized linguistic databases;
67,9 % - Intelligent data analysis
46,4 % - Methods for processing the text corpus;
35,7 % - Methods of processing the audio case;
32,1 % - Programming in Python
39,3 % - Application packages for processing speech signals;
53,6 % - Statistical methods of natural language processing;
46,4 % - Professional internship;
3,6 % - Programming in the Java language.

9. In your opinion, what forms and methods of instruction are preferred for the acquisition of these specific competences?
60,7 % - Theoretical;
67,9 % - Case studies;
53,6 % - Technical;
78,6 % - Practical internships;
35,7 % - Group work;
28,6 % - Digital lessons;
3,6 % - Other

10. As an expert, do you consider the need for professional training in Computational Linguistics?
If so, what are your needs?
Specialization in linguistics makes it easier to understand the principles of constructing and using formal controls for computer systems. For example, it simplifies the understanding of the paradigms of different programming languages, the transition between them and rapid adaptation; if necessary, the creation of their own language components with specialized syntax. Specialists in this field (with an applied bias in programming and system architecture development) are very much in demand.
variational phonetic transcription
Syntactic and morphological analysis of the text
basics of data mining

The application of mathematical methods in linguistic studies, for example, WSM (Word-Space Model)
language translation, text to gesture traslation, NLP
programming languages
as an expert should learn how to build internal algorithms of speech synthesis.
Theoretical basis
Work with modern linguistic information resources and expert computer systems

11. Are you personally interested in obtaining professional training in Computational Linguistics?
If so, what would you like to learn? In what field?
Development of programs for speech recognition
Phonetics of the tongue
In the field of building a computer model of natural language (Kazakh)
Digitalization of speech
programming languages
IT
Text recognising
In the field of education.

12. Are you interested in taking interns of the specialty «Computational Linguistics»?
If yes, what type of internship
35,7 % - Research;
71,4 % - Practical internship;
14,3 % - Functional internship with subsequent employment;
7,1 % - None.

If so, for what period would you accept the trainees?
42,9 % - from 1 to 2 weeks;
28,6 % - 1 month;
14,3 % - 3 months;
7,1 % - more than 3 months;
7,1 % by mutual agreement;
7,1 % - No.

Choice of disciplines for the educational program on Computational Linguistics

Table 2 lists the disciplines on Computational Linguistics of various universities.

Table 2. List of disciplines in Computational Linguistics

University	Disciplines
RGGU Russian State University for the Humanities	Introduction to fundamental linguistics Typology, comparativistics, areal linguistics Modern syntactic theories Case and experimental methods in semantics Introduction to Computational Linguistics Computer Sociolinguistics Mathematical Foundations of Linguistics Statistical models in linguistics Methods of artificial intelligence in Computational Linguistics Programming of linguistic tasks Linguistic annotation / markup Specialized linguistic databases Methods for evaluating AOT systems Models and methods of Computational Linguistics Classification methods and machine learning Linguistic basis of machine translation Computer Parsing Analysis of oral speech
SPbSU St. Petersburg University	Methods and models of ontological engineering Methods of knowledge engineering in humanitarian research Text understanding systems Text analysis models and their software implementation Statistical methods in language engineering Hull methods in language engineering Linguistics of the text and theory of verbal communication Languages and standards for describing information resources Expert systems and methods of inductive generalizations Methods of decision support Methods of software implementation of intelligent information technologies Mathematical modeling in data processing technologies Methodology and technology of designing information systems Information Society and Problems of Applied Informatics Business English Philosophical problems of science and technology

MPhTI Moscow University of Physics and Technology	Mathematical Foundations of Linguistics Statistical models in linguistics Introduction to fundamental linguistics Typology, comparativistics, areal linguistics Russian corpus grammar Introduction to Computational Linguistics Computer Sociolinguistics Modern syntactic theories Typology of grammatical categories English for professional communication Models and methods of Computational Linguistics Data structures and basic algorithms The main algorithms of linguistic analysis Analysis of oral speech. Corpus linguistics: building and using enclosures Classification methods and machine learning Computer models of discourse Linguistic basis of machine translation Formal models and resources of world languages Linguistic annotation / markup Methods for evaluating AOT systems Computer parsing. Methods of artificial intelligence in Computational Linguistics Application Packages for Linguistic Studies Specialized linguistic databases Linguistic support of the tasks of document analysis Automatic assessment of the complexity of texts

HSE High School of Economics Computational Linguistics	Linguistic data: quantitative analysis and visualization Introduction to Linguistics Mathematics Formal models in linguistics Functional and cognitive models in linguistics Computational Linguistics Programming (Python) Analysis of linguistic data: quantitative methods and visualization (taught in English) Mathematical foundations of Computational Linguistics Machine learning Experimental Linguistics Database Ontologies and semantic technologies Digital Humanitarian Technologies: Resources, Tools, Case Studies Designing of linguistic resources and systems
HSE High School of Economics Theory of Language	Introduction to Computational Linguistics Formal models in linguistics Functional and cognitive models in linguistics Analysis of linguistic data: quantitative methods and visualization Computational Linguistics Programming (Python) Experimental Linguistics Anthropology Theoretical models in the description of the language Comparative-historical linguistics Russian Studies Sociolinguistics Typology Machine translate Humanities in the Digital Age Digital Humanitarian Technologies: Resources, Tools, Case Studies
ITMO (St. Petersburg) St. Petersburg National Research University of Information Technologies, Mechanics and Optics	**Information Technology:** • System analysis and modeling of information processes and systems; • Designing information systems; • Organization of design and development of distributed systems software; • Organization of software design and development for embedded systems; • Software testing; • Quality management software development. **Speech technologies:** • Digital signal processing; • Digital processing of speech signals; • Mathematical modeling and decision theory; • Pattern recognition; • Recognition and synthesis of speech; • Recognition of the speaker (speaking by voice); • Multimodal Biometric Systems.
University of Oxford http://www.ox.ac.uk	Analysis of functional and structural data images of the brain. Physiological neuroimaging. Brain disorders. Diffusion of the image. Speech and the brain. Visualization. Neurodegeneration. Cognition. Psychiatry.

University of California, Los Angeles (UCLA) http://www.ucla.edu	Phonetics. Phonology. Syntax. Semantics. Psycholinguistics. Matlingvistics. Historical linguistics. African, Indian languages.
Harvard University https://www.harvard.edu	Fundamental studies of the speech apparatus and speech functions. Clinical studies of human voice and speech abnormalities. Mechanics, biophysics, physiology and / or molecular biology of the middle and inner ear. Acquired or congenital abnormalities of the mechanisms of hearing. Neurophysiological or modeling approaches in the study of nerve cells and circuits underlying auditory processing. Neurovisual studies of the mechanisms of tinnitus. Cognitive neurobiology of language signal processing. Design, development and improvement of the hardware and software system for hearing aids, ear implants, vestibular prostheses or algorithms for automatic speech recognition.
Cambridge University https://www.cam.ac.uk	Acoustic modeling (statistical models). Fundamental research in machine learning. Optimize dialogue using reinforcement learning. Recognition on large dictionaries. Pattern recognition. Speech recognition on mobile devices. Dictator independence and noise cancellation. Dialog systems and VoiceXML. Statistical language modeling. Statistical machine translation. Processing and transcription of recognized speech.
Carnegie Mellon University https://www.cmu.edu	User Interface Software Cognitive models. Speech recognition. Understanding of natural language. Computer graphics. Handwriting recognition. Visualization of data, visual design, multimedia. Computer support for teamwork. Computer music and theatrical skill. Social technologies.
Johns Hopkins University https://www.jhu.edu	Language modeling. Natural language processing. Neural treatment. Acoustic processing. The theory of optimization. Language Entry

program in three areas: "Linguistics", "Speech technologies", "Text technologies (word processing)".
Preparation of masters in the direction of "Linguistics" requires: a deep study of the fundamental foundations of linguistics with an emphasis on the methods of creating operational formal models of the language system, the adequate complexity of such tasks of natural language processing, as recognition and synthesis of speech and text, semantic analysis and understanding of text and speech.
Preparation of masters in the direction of "Speech Technology", requires in-depth study Methods of transcription of sounds and speech, Methods of Speech Recognition, Methods of speech synthesis, Methods of transcription of sounds and speech, Methods of Speech Recognition , Methods of speech synthesis
Preparation of masters in the direction of "Text Technology" requires Methods and tools for creating text corpora
Methods and tools for creating audio corpora, Syntax analysis of texts, Morphological analysis of texts, Semantic analysis of texts

It should be noted that any educational program consists of mandatory disciplines (mandatory part) and elective disciplines (the variable part).
To ensure the training of specialists in Computational Linguistics in the three areas, the following disciplines were selected from Table 2:
1. **Mandatory part**: History and philosophy of science, Foreign language (professional), Pedagogics, Psychology, Software development technology
2. **Variational part**:
Linguistics: Ontology design tools, Tools for processing visual data, Tools for processing audio data , Tools creating thesauri, Semantic Search Tools.
Speech technologies: Methods of transcription of sounds and speech, Methods of Speech Recognition, Methods of speech synthesis, Methods of transcription of sounds and speech, Methods of Speech Recognition , Methods of speech synthesis.
Text technology: Methods and tools for creating text corpora
Methods and tools for creating audio corpora, Syntax analysis of texts, Morphological analysis of texts, Semantic analysis of texts.

Educational program "Computational Linguistics»

Table 3 lists the competencies for the educational program «Computational Linguistics»

Learning Outcomes	**Short description**
Knowledge of subjects	IT, Computational Linguistics
Specific skills	C1. The ability to build and solve mathematical models in accordance with the direction of training and specialization C2. The ability to apply modern programming languages and data manipulation languages, operating systems, software packages, etc. in research and application activities. C3. The ability to publicly present the results of professional activity (including the use of information technology) C4. Ability to apply existing tools and linguistic resources for the development of applied text processing systems in natural language. C5. Knowledge of the basic levels of analysis and synthesis of text in natural language, existing models of statistical, morphological and syntactic analysis of texts and their application in typical software applications for text processing; C6. Knowledge of the principles of constructing various linguistic resources, including the corpus of texts, terminological dictionaries, thesauruses, ontologies; C7. Understanding the essential differences between natural languages and artificial models of computer models of natural language; C8. The ability to carry out the processing of linguistic data by modern means C9.Knowledge of the algorithms of primary processes for automatic processing of text and speech C10. The ability to understand which language structures lie behind a particular local task, what are the linguistic and extralinguistic properties of texts used to solve problems. C11. knack to use statistical methods for analyzing language data and tools for their visualization C12.Knowledge of various methods of mathematical generalization of the results of linguistic research. C13. knack to design specialized linguistic databases
	C14. knack to correctly use the results of mathematical generalization and use the data obtained to solve the tasks in the research and qualification work. C15. Knowledge of the basic approaches to digital modeling of humanitarian data in the field of history and literature, the main trends in the development of this field of knowledge C16. knack to navigate in computer tools of linguistic analysis and computer linguistic resources C17. knack to evaluate the complexity of different solutions and the thresholds of acceptable errors C18. Knowledge of the modern electronic environment C19. knack to program prototypes and decision models; prepare the necessary linguistic resources. C20. The ability to design a chain of processing of language data and interpret the results of automatic processing of linguistic data. C21. The ability to understand the current state of Computational Linguistics and information technology C22. The ability to analyze, compare and critically evaluate different linguistic directions, theories and hypotheses C23. The ability to conduct independent research and obtain new scientific results in the field of language theory, linguistics of specific languages, applied and Computational Linguistics

Interdisciplinary skills	D1 - Work in an interdisciplinary team, the ability to interact with experts in other subject areas. D2 - Work in an international context. D3 - Comparison, analysis and interpretation of complex experimental information and the formulation of conclusions. D4 - Solving the theoretical and practical problems of computer tools and software in a variety of contexts and the ability to establish relationships between problems and basic principles. D5 - Solving a wide range of known theoretical and practical problems of computer tools and software and implementing solutions to implicit and unresolved problems. D6 - development of large-scale computing experiments in applied fields. D7 - Forecast of possible weaknesses and risks of performing research. D8 - Organization and planning of professional, scientific and scientific-pedagogical activities, as well as the activities of the team. D9 - Critical thinking, criticism and self-criticism. D10 - Carrying out scientific research and working as a team leader.

Distribution of disciplines by educational modules
Master=2 ans (120 ECTS)
KZ : Master = 2 ans 1 year of studies=60 ECTS

UE 1. Methodology of master's training **(Sociocultural aspects of Computational Linguistics)**
- History and philosophy of science
- Foreign language (professional)
- Pedagogics
- Psychology

UE 2. Software development technology

UE 3. Methods of natural language processing
- Statistical methods in Natural Language Processing
- Mathematical Foundations of Computational Linguistics

UE 4. Methods and tools for the creation of language corpora
- Methods and tools for creating text corpora
- Methods and tools for creating audio corpora

UE 5. Languages for symbol processing
- Programming in Python
- Programming in Prolog

UE 6.Digital Signal Processing
- Methods of digital processing of speech signals
- Software for processing speech signals

UE 7. Tools for processing natural languages
- Ontology design tools
- Tools for processing visual data
- Tools for processing audio data

UE 8. Text recognition
- Syntax analysis of texts
- Morphological analysis of texts
- Semantic analysis of texts

UE 9.Speech recognition
- Methods of transcription of sounds and speech
- Methods of Speech Recognition
- Methods of speech synthesis

UE 10.Big Data processing and storing
- Specialized linguistic databases
- Data Mining

UE 11. Search Tools
- Tools creating thesauri
- Semantic Search Tools

UE 12.Creating of text corpus
- Sentiment analysis of natural language texts
- Method for processing of text corpora

UE 13.Creating of audio corpus
- Synthesis of speech analysis of natural language
- Methods for processing of audio corpora

UE 14. Practices
- Scientific-research work of graduate students (master)
- Teaching practice
- Research internship

Decomposition of the curriculum in semesters

Master = 2 year 120 ECTS
1 year of study =30 ECTS

Year	Semester	Name of semester (*)	UE
Year 1	S1		UE1 History and philosophy of science UE2Foreign language (professional) UE3Statistical methods in Natural Language Processing UE4Mathematical Foundations of Computational Linguistics UE5 Methods and tools for creating text corpora UE6 Methods and tools for creating audio corpora UE7Programming in Python UE8 Programming in Prolog UE9Methods of digital processing of speech signals UE10Software for processing speech signals UE11Scientific-research work of graduate students (master 1)
	S2		UE12 Pedagogics UE13 Psychology UE14 Software development technolog UE15 Ontology design tools UE16 Tools for processing visual data UE17 Tools for processing audio data UE18Syntax analysis of texts UE19Morphological analysis of texts UE20Semantic analysis of texts UE21 Methods of transcription of sounds and speech UE22Methods of Speech Recognition UE23Methods of speech synthesis UE24Scientific-research work of graduate students (including scientific internships)
Year 2	S3		UE25Specialized linguistic databases UE26Data Mining UE27Tools creating thesauri UE28Semantic Search Tools UE29 Sentiment analysis of natural language texts UE30 Method for processing of text corpora UE31Synthesis of speech analysis of natural language UE32Methods for processing of audio corpora UE33Scientific-research work of graduate students UE34Teaching practice
	S4	Practice	UE35Scientific-research work of graduate students UE36Research internship

(*) thème général du semestre d'un point de vue pédagogique

Description of UE

Semester 1 UE (1 semester = 30 ECTS)

UE	Objectif	Modules	ECTS	Lectures	TP	TL	W pers.	Total
UE1	History and philosophy of science	Methodology of master's degree students training	3	15	15		60	90
UE2	Foreign language (professional)		3		30		60	90
UE3	Statistical methods in Natural Language Processing	Methods of natural language processing	5	15	30		90	135
UE4	Mathematical Foundations of Computational Linguistics		5	15	30		90	135
	course by choice 1*							
UE5	Tools for creating text corpora	Tools for the creation of language corpora	5	15	30		90	135
UE6	Tools for creating audio corpora		5	15	30		90	135
UE7	1.1 Programming in Python	Languages for symbol processing	5	15	30		90	135
UE8	1.2 Programming in Prolog		5	15	30		90	135
UE9	1.3 Methods of digital processing of speech signals	Digital Signal Processing	5	15	30		90	135
UE10	1.4 Software for processing speech signals		5	15	30		90	135
UE11	Scientific-research work of graduate students (master)		4					120

*- the student has the right to choose two of the 4 courses. Each course is 5ECTS (135 hours)

Conventions:
TP: Practical work
TL: Laboratory work
Pers. work: Independent work (libraries, at home, in practice, etc.)
ECTS: European credit transfer system

Semester 2 UE (2 semester = 30 ECTS)

UE	Objectives	Modules	ECTS	Lectures	TP	TL	Pers. work	Total
UE12	Pedagogics	Methodology of master's degree students training	3	15	15		60	90
UE13	Psychology		3	15	15		60	90
UE14	Software development technology		3	15	15		60	90
	Course by choice 2**							
UE15	Ontology design tools	Tools for processing natural languages	5	15	30		90	135
UE16	Tools for processing visual data		5	15	30		90	135
UE17	Tools for processing audio data		3	15	15		60	90
UE18	Syntax analysis of texts	Text recognition	5	15	30		90	135
UE19	Morphological analysis of texts		5	15	30		90	135
UE20	Semantic analysis of texts		3	15	15		60	90
UE21	Methods of transcription of sounds and speech	Speech recognition	5	15	30		90	135
UE22	Methods of Speech Recognition		5	15	30		90	135
UE23	Methods of speech synthesis		3	15	15		60	90
UE24	Scientific-research work of graduate students (including scientific internships)		8					240

** - the student has the right to choose three out of 6 courses. Each course - 3 credits (90 hours) or 5 credits (135 hours)

Semester 3 UE (3 semester = 30 ECTS)

UE	Objectives	Modules	ECTS	Lectures	TP	TL	Pers. work	Total
UE25	Specialized linguistic databases	Big Data processing and storing	5	15	30		90	135
UE26	Data Mining		4	15	30		90	135
	course by choice3* **							
UE27	Tools creating thesauri	Search Tools	5	15	30		90	135
UE28	Semantic Search Tools		5	15	30		90	135
UE29	Sentiment analysis of natural language texts	Creating of text corpus	5	15	30		90	135
UE30	Method for processing of text corpora		5	15	30		90	135
UE31	Synthesis of speech analysis of natural language	Creating of audio corpus	5	15	30		90	135
UE32	Methods for processing of audio corpora		5	15	30		90	135
UE33	Scientific-research work of graduate students		8					240
UE34	Teaching practice		3					90

***- the student has the right to choose two of the 4 courses. Each course - 5 credits (135 hours)

Semester4 UE (4semester = 30 ECTS)

UE	Objectives	Modules	ECTS	Lectures	TP	TL	Pers. work	Total
UE35	Scientific-research work of graduate students		8					240
UE36	Research internship		12					360
UE37	Complex exam		3					105
UE38	Writing and defense of Master's degree thesis		7					315